新兴研究主题在演化路径上的关键时间点研究*

■ 许海云[1] 张慧玲[2] 武华维[3] 刘自强[4]
[1] 山东理工大学管理学院 淄博 255000 [2] 太原市图书馆 太原 030024
[3] 西北师范大学档案馆 兰州 730070 [4] 南京师范大学新闻与传播学院 南京 210023

摘 要：[目的/意义]探讨不同关键时间点对新兴研究主题影响力的不同表征。[方法/过程]首先，综述当前拐点时间的应用场景及获取方法，并根据知识扩散中网络节点增长机制与特征构建新兴研究主题在创新演化路径上的拐点识别方法。其次，对比分析首次出现时间、平均时间和拐点时间的差异，探讨新兴研究主题产生影响力的最早时间点。最后，以干细胞研究主题为实证领域，分析不同关键时间点对新兴研究主题影响力的不同表征能力。[结果/结论]拐点时间可以比平均值时间提前识别有影响力的主题。首次出现时间、平均时间和拐点时间在主题发展路径中意义区别显著，新兴研究主题在创新路径中分布时间的确定需要综合3种不同类型的关键时间点。

关键词：演化路径 创新路径 拐点时间 新兴研究主题 曲线拟合
分类号：G250.2
DOI：10.13266/j.issn.0252-3116.2021.08.006

1 引言

新兴研究主题识别作为预测科技创新研究的重要手段，也指示了科技创新研究的发展方向。当前主要的科技创新型国家和地区不仅纷纷资助新兴研究，还资助利用科学计量方法评估受资助项目的研究效果。例如欧洲研究委员会（European Research Council，ERC）自2000年开始资助两个项目：前沿研究识别文献计量模型的开发和论证（Development and Verification of a Bibliometric model for the Identification of Frontier Research，DBF）、ERC支持的新兴研究领域及其评估（Emerging Research Areas and their Coverage by ERC-supported projects，ERACEP）[1]。其中，ERACEP项目分析了2009年ERC支持的研究涵盖的新兴领域情况[2]。

美国情报高级研究计划局（The Intelligence Advanced Research Projects Activity，IARPA）2011年建立的科学博览的预见与理解计划（Foresight and Understanding from Scientific Exposition，FUSE）以产生变革性创新为使命，关注"高风险/高回报"项目，不追求短期利益，尤其重视重大新兴研究主题的早期识别研究[3]。前瞻性识别新兴研究主题，对科技管理人员部署创新战略和优化资源配置至关重要。因此，新兴研究主题一直是重要的学术研究对象，当前，新兴主题和新兴研究越来越多的出现在新闻出版物和学术期刊中。与此同时，科学计量学已经开发出多种方法来监测科技创新中的新兴研究主题[3]，其中演化路径是研究新兴研究主题的重要途径之一。

新兴研究主题的演化路径是指新兴主题的发展脉络，反映了技术创新的产生、扩散和演化过程。引文分析是常用的科技演化路径获取方法，领域主题基于引文数据的相似性，按照时间轴聚类后可获得主题演化路径[4]。通过新兴主题的演化路径，可以深入了解主题演化的背景及突破性创新的机制，辅助识别并预测突破性创新。除此之外，演化路径还可以辅助判断新兴研究主题的新颖性、增长/持久性、连贯性等不易量化的特征，通过可视化方式来分析不确定性和模糊性指标[5]。

一个新兴研究主题在其生命周期内可能多次出现在创新演化路径上，并且可能出现在不同的时间间隔

* 本文系国家自然科学基金项目"基于科学-技术主题关联分析的创新演化路径识别方法研究"（项目编号：71704170）和中国科学院十三五信息化项目"面向干细胞领域知识发现的科研信息化应用"（项目编号 XXH13506）研究成果之一。
作者简介：许海云（ORCID：0000-0002-7453-3331），教授，博士，E-mail：xuhaiyunneomo@gmail.com；张慧玲（ORCID：0000-0001-5155-6357），馆员，硕士；武华维（ORCID：0000-0001-6969-407X），馆员，博士；刘自强（ORCID：0000-0003-1814-8655），讲师，博士。
收稿日期：2020-09-03 修回日期：2020-11-30 本文起止页码：51-64 本文责任编辑：王传清

中,最终会形成一个主题的时间分布。这些主题的时间分布并非均匀和等价的,而是存在少数重要的“关键时间点”。关键时间点是指主题产生、发展、演化和突变等关键时间点。在演化路径上不同类型的时间点对标识新兴研究主题的生命周期有何区别? 应该如何运用不同的时间点提高对新兴研究主题的监测和趋势追踪能力? 当前,科学文献累积值的平均时间值通常被用作科学技术发展路线图中的关键时间点,但实际上论文发表与专利申请的趋势并非线性累积,而是近似遵循指数发展的规律,呈现偏斜态分布。因此,如果在主题演化路径的时间轴上选择平均时间作为其关键时间点将导致主题识别的滞后。与此同时,最早出版年或申请年可能仅能表示零星的研究,尚未形成较大的影响。

鉴于此,笔者分析了 3 种类型的关键时间点对表征新兴主题发展影响力的差异,尤其是对新兴技术早期识别具有重要作用的拐点时间,同时探讨新兴研究主题最有影响力的关键时间点,以便更清晰地显示新兴研究主题在演化路径上的关键时间点,在此基础上,对比分析其他关键时间点(首次出现时间、平均时间)与拐点时间的差异。本研究旨在改进主题演化路径的准确性,以期提高主题演化路径在预测新兴研究主题识别中的准确性。

2 相关研究

分析关键时间点对表征新兴主题发展影响力的差异,需要了解 3 个相关问题:新兴研究主题的演化特征、拐点时间在不同学科领域的应用、拐点时间的获取方法。笔者从这 3 个方面综述已有研究以期为本文提供借鉴。

2.1 新兴研究主题的演化曲线特征

科学文献传播网络的形成初期带有随机性特征,传播网络初期,有的时间段不出现新增节点,而有的时间段会出现多个新增节点,网络新增节点的增长规律与网络结构相关,其增速呈现先快后慢的总体趋势。随着传播网络规模的扩大,新加入的网络节点数目逐渐增多,而当学科领域发展到某一阶段,受限于研究人员的学识水平和专业领域的信息资源等,传播网络的增速将呈现放缓趋势,最终达到相对稳定的状态。

P. Young[6] 收集了许多代表增长行为的时间序列,并根据数据分布特征对其分类,之后尝试确定不同分布类型的最佳增长曲线预测模型,由此可以在拟合模型之前根据数据集的特征为特定的数据集选择最优增长曲线模型。有些研究主要集中在技术增长预测过程的阶段问题与产业发展曲线模型修正上。D. Kucharavy 等提出了基于 S 曲线的 3 个阶段的短期、中期和长期预测的区别[7],强调了 Logistic 增长模型的增长参数的定义[8],并且与 R. D. Guio[9] 一起讨论了基于 Logistic 生长函数的简单 Logistic、成分 Logistic 和 Logistic 替代模型在技术变化预测中的应用,有助于新兴技术的长期预测。P. Nguimkeu[10] 为了避免使用的增长曲线导致预测误差,对 Gompertz 和 Logistic 增长模型进行了简单模型的选择检验。N. Meade 等[11] 考虑了不同国家不同阶段的扩散创新,在模型参数中引入营销变量,并对模型进行了标准化处理,以实现扩散的持续性。

除此之外,技术增长模型在新兴技术领域的研究也逐渐成熟,T. Daim 等[12] 将文献计量学和专利分析与情景规划、增长曲线和类比等著名技术预测工具相结合,预测了燃料电池、食品安全和光存储 3 个新兴技术。M. Kyebambe 等[13] 为了验证所提出的监督学习自动标记数据方法的高效性,利用美国专利商标局提供的专利引用数据,使用专利引用方法从专利引用数据中得出预测因子,可以在新兴技术出现至少一年之前实现 70% 的新兴技术预测。

可见,当前对于新兴研究主题的演化曲线特征已经有诸多研究,演化曲线存在多种演化特征,针对不同的曲线特征又衍生出多种拟合与预测模型。

2.2 拐点时间的应用

不同的学科领域,拐点的内涵有所不同。从哲学角度看,拐点是解决矛盾和事物转变的关键点[14-15],即拐点是实现事物量变到质变、渐变到突变的点。数学上,定义拐点为平面曲线弯曲方向发生改变的转折点,若曲线图形在一点由凸转凹,或由凹转凸,即拐点是使切线穿越曲线的点。在拐点求解中,若该曲线图形的函数在某点的二阶导数为零或不存在,且二阶导数在该点两侧符号相反,则该点为函数的拐点[16-18]。在经济管理领域,拐点是事物发展曲线中具有重大意义的点,能够展望未来的不确定性和预警潜在风险。尽管各个领域拐点的具体含义有所差异,但普遍意义上拐点是指数据或事物增长与下降、兴盛与衰败的转折点,拐点前后演化趋势不一致[19]。

各个领域拐点的不同内涵体现为曲线拟合中自变量与因变量的不同。哲学领域中曲线拟合的自变量大多是上层建筑中的意识形态,因变量则是社会发展;数学领域多研究拐点测量的公式模型;经济管理领域关

注市场与金融方面的拐点，以达到预测走势的目的，通常价格、产量、销售额等被定义为自变量，经济发展水平与行业发展水平定义为因变量。也有部分专家学者将拐点应用于舆情预测等方面，丰富了拐点应用范围。代表性研究如下：

在经济管理领域，专家学者多利用拐点探测经济指标的发展趋势。例如，“刘易斯拐点”[20－22]利用农业劳动的边际生产率预测拐点，由此判断劳动力供给趋势。“库兹涅茨拐点”[23－25]基于收入分配与经济增长之间的关系，判断收入分配差距。“蒙格斯拐点”则是测算边际收益与边际成本的均衡点[26]。

在拐点的应用过程中，不少专家学者结合领域特点，面向具体应用场景，进一步丰富了拐点的类型。石陆魁等[27]在预测股价范围的研究中，利用时间－股价(y)的关系，定义了上拐点与下拐点，满足$(y_n - y_{n-1})/y_n > R$且$(y_n - y_{n+1})/y_n > R$，为上拐点，满足$(y_{n-1} - y_n)/y_n > R$且$(y_{n+1} - y_n)/y_n > R$，则为下拐点，其中R为设定阈值。郑步青等[19]结合网络舆情预测，将拐点分为阶段拐点与极值拐点，阶段拐点将舆情演化分为潜伏、爆发和消亡3个阶段。

由此可见，拐点时间在不同学科领域的内涵不同，不同领域对拐点时间的应用也存在差异。但不同领域对“拐点时间可以表征研究对象的重要转折时间点”比较认同。

2.3 基于曲线拟合的拐点获取方法

近年来，贝叶斯面板VAR模型[28－30]和隐马尔可夫模型[31－33]等方法开始应用于拐点预测，F. Canova等[34]提出基于VAR贝叶斯面板的转折点预测的方法，用于7国集团中产出增长率的拐点预测问题。M. Billio等[35]将VAR模型和马尔可夫模型相结合，生成用于业务周期分析的拐点预测方法；L. Koskinen等[36]采用隐马尔可夫模型，对瑞典和美国经济的拐点进行预测。之后，R. F. Pelàez等[37]提出一个Logit模型能够更准确地预测商业周期拐点。N. Kulendran和K. K. F. Wong[38]将Logit模型与Probit模型结合预测经济周期的拐点。

除此之外，时间序列预测模型中的ARIMA模型[39－40]与混沌时间序列分析[41－43]也被广泛应用于拐点预测之中。经济学家发现拐点的改变与增长周期之间存在着一定关联，但时间序列模型不能显著呈现出增长趋势与增长周期，因此，增长曲线模型(Growth Function)[38,44－45]也常被应用于拐点的研究。其中Logistic曲线分析在预测中的一个典型应用是M. King Hubbert对美国石油工业储量和产量的分析。1962年，M. King Hubbert将测量国内石油累积探明储量和累积产量的Logistic曲线结合起来，预测了1968年美国石油产量的峰值，比实际峰值提前了两年[46]。此后，一些学者纷纷开始运用Logistic曲线进行拐点分析的相关研究。詹新宇利用Logistic曲线方程预测广西产业结构服务化的拐点，发现使用增长曲线模型预测拐点的准确性较高[45]。

随着知识关联的不断深入，灰色预测[47－49]与神经网络[50－52]开启了拐点预测方法的新方向。田红丽等[47]运用相似灰色模型对股价拐点进行预测，解决了灰色预测模型在振荡序列中预测拐点精度不高的问题。O. Claveria等[50]通过比较多层感知器网络、径向基函数网络和埃尔曼网络3种不同的人工神经网络技术在游客需求应用的拐点预测，发现多层感知器网络在复杂预测中的精确度较高，若增加维度可以提高拐点预测的精确度。

由于产品和服务都需要一个生命周期，这个生命周期包括萌芽、发展、成熟和衰退，确定产品在其生命周期中的位置是预测拐点的另一种方法[53]。P. C. GUO等根据生命周期理论，发现当物流指数超过均衡值时，物流业会对经济产生放大效应并迎来物流业发展的拐点[54]。P. N. Golder等[55]发现部分产品销售曲线的急剧增长大多出现在新产品的引入阶段到成长阶段，并将这个现象称为“起飞(Take Off)”，将过渡点称为“起飞点”。

不同的拐点识别与预测模型，均旨在提高特定领域拐点预测的准确性。缘于不同领域甚至相同领域的不同对象有不同的生命周期曲线，因此没有哪种方法是最优的，需要根据该领域数据分布特征具体分析。

3 创新路径关键时间点与获取方法

3.1 主题分布的拐点时间

新兴研究主题是创新路径上可以持续发展的主题。在这一发展过程中，新兴研究主题在创新路径上的分布与科学发现及技术创新的规律一致，并不是时间轴上的均匀分布。正如科技创新的突破往往呈现集群式突破特征，创新路径上主题包含的产出会显示出多种模式，如逐渐增加、逐渐减少、先升后降、先降后升，以及其他组合形式。因此，如何展示主题在时间轴上的动态变化值得深入分析。

笔者将主题中包含的科学文献的数量作为对主题的关注度和影响力的标志。通常，在取得较大进展或

突破时，研究主题将吸引从事相同或密切相关研究的科学家的注意甚至加入，并且随着时间的推移该主题进展得到更广泛的扩散，得到更多关注，表现为科学出版物的数量也会迅速增加并最终趋于平稳。这个过程可以用增长函数曲线表示[56]。从理论上讲，拐点可以在较早的时间点捕获有影响力的新兴研究主题。图 1 为创新主题的生长曲线分段示意图。图 1 中星号（★）标记的时间点是论文发表数量正增长率发生变化的点，标志着增长率转为下降的时刻。拐点附近的主题处于快速增长时期，在拐点附近，主题已经发表了一定数量的科学文献但尚未达到最大数量，且处于快速增长区间。当该主题的出版物数量结束快速增长进入平稳期时，则表征着该主题已成为一个研究热点，论文发表量已接近最大数量，而此时，主题已过了最有创新影响力的时刻。

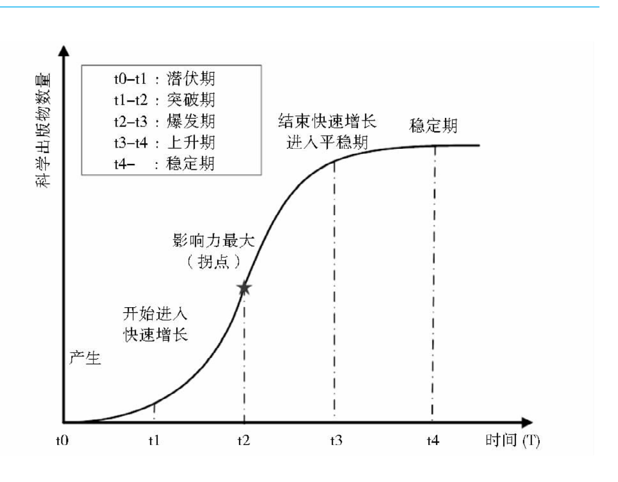


图 1　创新主题的生长曲线分段示意

3.2　分析方法与流程

首先，识别新兴研究主题，根据新兴研究主题的识别指标（主题的新颖度、增长率、潜在社会影响力、持续性/连贯性、不确定性），筛选出新兴研究主题作为对比分析对象。其次，绘制新兴研究主题论文发表时间分布的散点图，进行曲线拟合获得 3 种关键时间点（拐点时间、首次出现时间、平均时间）。再次，绘制拐点时间在演化曲线图中的坐标标签，并详细分析不同拐点类型。最后，对 3 种关键时间进行对比分析。分析方法模型与分析流程见图 2。

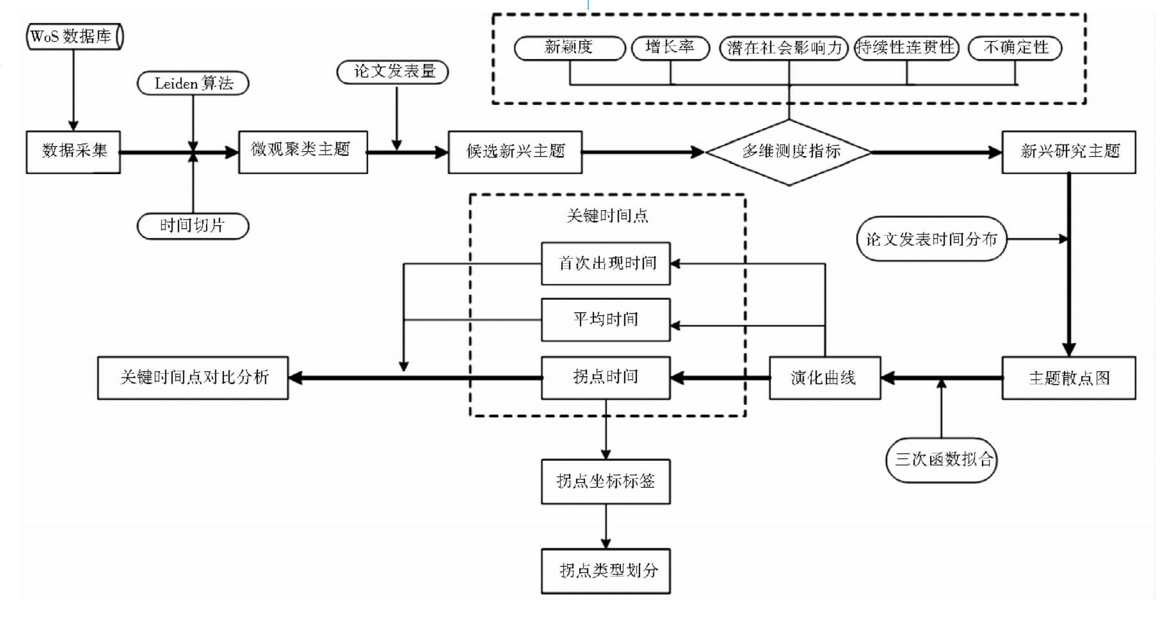


图 2　方法模型与分析流程

3.2.1　获取新兴研究主题

新兴研究主题是指新颖度高、发展相对快速，具有一定程度的连贯性和较大的科学影响力的研究主题，并有可能对社会经济领域产生重大影响。笔者基于以往研究开展进一步探索，先前研究中详细论述了如何识别新兴研究主题的多维度科学计量指标，包括主题的新颖度、主题的增长率、主题发展的潜在经济社会影响力、主题发展的持续性和连贯性、未来发展的不确定性分析[5,57]。

3.2.2　识别拐点时间

主题的相关出版物数量代表着该主题在学术领域内的影响。本文将干细胞研究的相关主题作为研究对象，根据每个主题的年论文发表量，绘制每个主题数据的散点图，形成该主题的研究发展路径。之后，通过曲线拟合，计算研究发展路径中的拐点，并针对新兴研究主题与非新兴研究主题拐点特征与效用进行对比分析。

笔者尝试利用生长函数(Logistic Function)曲线进行散点的曲线拟合，但效果不理想。通过尝试发现一元三次函数可以更好地拟合本研究的散点分布，因此，笔者采用三次函数拟合散点图，在图像中绘制函数曲线，输出拟合的函数。

其中，创新路径上拐点获取方法如下：

第一步，从原始数据文件中提取样本数据，统计每一个主题年度(x)所发表的论文篇数(y)，据此绘制每一个主题数据的散点图。

第二步，选择拟合函数拟合第一步中的散点图，可视化拟合曲线并输出拟合函数。

第三步，根据拟合曲线计算拐点值，并在图像中绘制拐点位置。

3.2.3 分析拐点类型

依据拐点时间出现的区间范围，将拐点类型分为4类：拐点出现在分析时间区间内、拐点出现在分析时间区间之前、拐点出现在分析时间区间之后、拐点未在分析时间区间内出现。

3.2.4 三种关键时间点的对比分析

为了进一步证实拐点在预测主题发展趋势方面的效果，笔者对比拐点在预测分析新兴研究主题与非新兴研究主题中的差异，并区分首次出现时间、平均时间以及拐点时间在主题预测分析中的功能。拐点与平均时间、最早时间计算公式如表1所示：

表1 拐点与平均时间、最早时间公式

特征时间	计算公式	变量解释	模型意义
拐点时间	$f(x)=ax^3+bx^2+cx+d$	a、b、c、d为参数	任何一个三次函数都存在一个拐点，所以当a≠0的时候，在定义域内都存在拐点
平均时间	$\bar{t}=\frac{1}{n}\sum_{i=1}^{n}t_i$	t为时间，n为时间点的总个数，i为起始的时间点	表明各时间点相对集中较多的中心位置
最初时间	$t=t_i,\ i=1$	t为时间，i为起始的时间点	表明最初开始的时间

4 实证分析

4.1 数据集与统计分析

干细胞是一类具有自我更新和多向分化能力的细胞，是生物医学领域的重要研究对象，因其在疾病治疗和再生医学方面的重要价值和巨大发展前景，干细胞在生命科学和医学研究中引起了关注。笔者选择干细胞领域作为本文方法的实证领域，选用Web of Science(WoS)获取分析数据，检索日期为2018年10月20日，共获取了干细胞研究相关的422 101篇研究论文。

具体检索策略为：TS =(("stem cells" OR "stem cell") NOT ("stem cellulose" or "stem Cellular" or "cello" or "cellar" or "cellphone")) OR TS =((ESC or ASC or iPS or PGC or MSC or CSC or LSC or TSC or ADSC or HSC) near (cells OR cell)) OR TS =((("totipotent" or "pluripotent" or "multipotent" or "unipotent" or "progenitor" or "precursor") NOT ("non-pluripotent")) ADJ (cells OR cell))) NOT TS =("fuel cell" or "in-plane switching" or "Intrusion Prevention System")。

社区识别是分析大型复杂网络结构的常用方法，本文通过引文网络识别有意义的研究主题。V. A. Traag等开发了Leiden算法[58]，该算法能发现更好的社区划分，提高了社区之间的有效连接且运行速度更快。在Leiden聚类算法下，WoS数据库的所有论文被分为3个层级，宏观层有22个主题，中观层有868个主题，微观层有4 047个主题。笔者基于Leiden算法采用了微观层主题分类，并在此基础上进一步分析新兴研究主题。

为了避免数据库的扩大或缩小等随机波动的影响，笔者采用平滑的年度发表量，选取2004－2018年为研究时间段，以每5个相邻年份为一组设置了11个时间切片。在研究时间段内，论文的发表数量呈指数增长(见图3)。422 101条记录的微观主题粒度分类涉及不同时间切片过程中的变化，如2004－2008年有1 584个微观主题，2013－2017有2 058个微观主题[58](见图3)。

4.2 识别主题和可视化分析

干细胞研究主题的分布见图4，节点的大小与干细胞研究的论文数量成比例，节点越大表示该主题的论文数量越多；主题之间的距离越小，表明研究内容越相近。图4节点上的数字不是所涉及的论文数量，而是该主题在整个WoS数据库中的微观层分类标签编号。干细胞研究相关主题主要位于“生物健康”区域且涉及主题数量较多。尽管Leiden分类算法[58]建立在整个WoS数据库的基础上，有助于更准确地发现全局主题，但却缺乏具体领域特异性。为得到更具针对性的干细胞主题，笔者进一步获取干细胞各主题在Leiden分类中的比例，并删除了仅包含少量干细胞论文的主题。

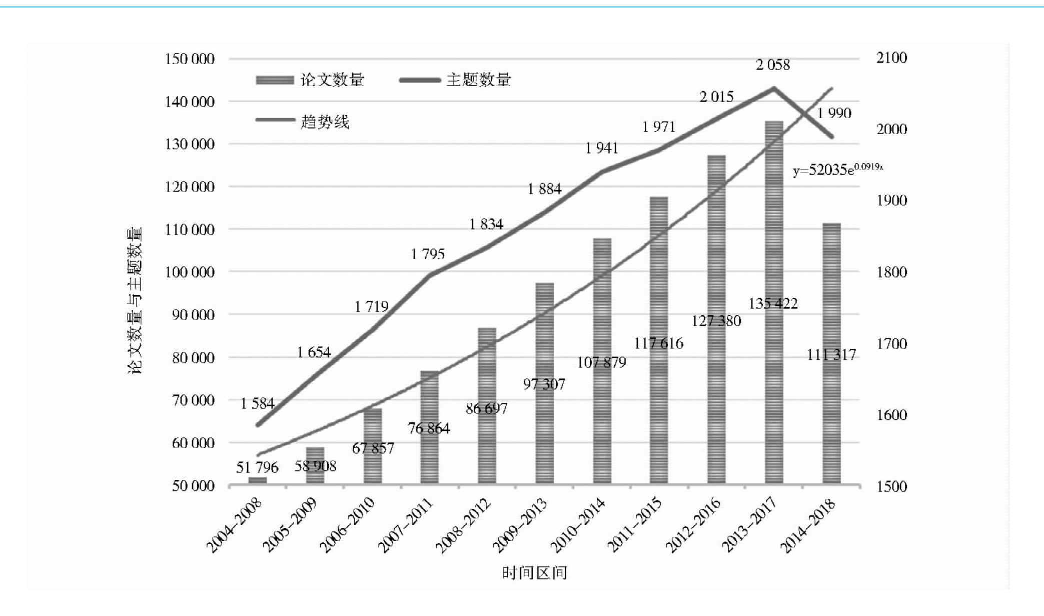


图 3　2004 年 –2018 年干细胞研究领域论文发表和主题数量趋势

注:检索时无法获取截至 2018 年的全部数据,因此图中的曲线末端下降不代表真实情况的数据

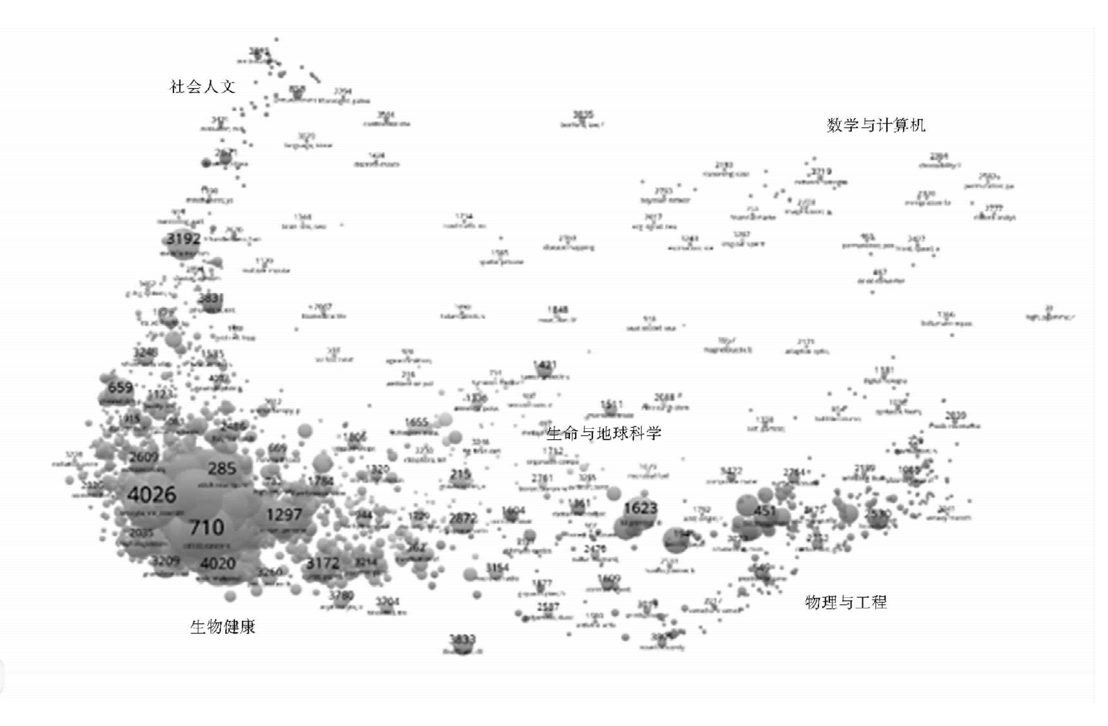


图 4　2004 –2018 年干细胞研究领域的主题分布

结合新兴研究主题的识别指标,即主题的新颖度和增长率对候选新兴研究主题进一步筛选。新颖度和增长率包含 4 个指标:论文数量、期刊数量、基金数量和作者数量的平均增长率。论文数量表征主题的研究热度,期刊数量表征该主题吸引的投稿数,基金数量表征主题的受资助情况反映国家的重视程度和投入力度,作者数量表征主题受学者的关注程度。新兴研究主题的筛选步骤如下[59]:

第一步,探究各个主题全阶段的增长态势。首先计算每个主题的论文数量在各阶段的增长率,再计算该主题论文数量在全阶段的平均增长率(i = 10),把握主题出现的时间和整体发展趋势。在此过程中,关注近期(2010 年之后)出现的主题,而对于近期消失的主题,本文认为其新颖性不足,予以排除,初步筛选出 54 个新兴技术主题。平均增长率公式为:

$$\overline{V}_i = \frac{1}{i}\sum_{t=1}^{t+i}\frac{N_{t+i}-N_t}{N_t} \qquad 公式(1)$$

其中 i 表示时间段,N 表示论文数量、期刊数量、

基金数量、作者数量，$\bar{V}^i$ 表示平均增长率。

第二步，计算与各个主题相关的期刊数量、基金数量、作者数量在全阶段的平均增长率（i＝10）。剔除平均增长率为空值的主题，空值表示该主题在期刊、基金或作者数量排名统计中没有进入前 50 名，表明该主题的研究热度或受关注度不够，故予以剔除。最后剔除 10 个主题，筛选得到 44 个新兴技术主题。

第三步，分析各个主题近期的增长趋势。选择后 5 个阶段作为分析时间窗（i＝4），计算剩余主题论文数量在后 5 个阶段的平均增长率。剔除后 5 个阶段平均增长率为负值的主题，负值表明该主题论文数量在近期呈递减趋势，共剔除 18 个主题。

最终，筛选出 26 个新兴研究主题，鉴于篇幅限制，表 2 列出了论文数量平均增长率（降序排列）最高的 10 个主题的主题标签和内容概述。

表 2　干细胞领域新兴研究主题（部分）

主题编号	主题标签（关键词）	主题概述
353	stem cell、cell、surface、mesenchymal stem、EXTRACELLULAR MATRIX、substrate、differentiation、tissue、hydrogel、biomaterial	Extracellular matrix could modulate stem cells’ fate
2276	stem cell、intestinal stem、expression、organoid、cancer、intestinal、COLORECTAL CANCER、crypt、model、Lgr5	Stem cells in intestinal epithelium
1460	stem cell、scaffold、tissue engineering、cell、extracellular matrix、tissue、mesenchymal stem、VITRO、regeneration、decellularized	Decellularized stem cell in tissue-regenerative medicine
142	DNA methylation、STEM CELL、epigenetic、gene、gene expression、EMBRYONIC STEM、expression、human、cell、development	DNA methylation in stem cells
60	stem cell、mesenchymal stem、ARTICULAR CARTILAGE、scaffold、tissue engineering、chondrocyte、chondrogenic differentiation、bone marrow、growth factor、cartilage repair	Stem cells provide sources for cartilage tissue engineering
921	stem cell、dental pulp、expression、periodontal ligament、mesenchymal stem、pulp stem、osteogenic differentiation、human dental、VITRO、growth factor	Human dental pulp stem/progenitor cells （hDPSCs） based therapy for bone tissue engineering
727	beta catenin、stem cell、Wntsignaling、expression、Wnt beta、signaling pathway、cell、catenin signaling、protein、gene	Regulation of Wntsignaling pathway in stem cells
1046	stem cell、EZH2、expression、cell、gene、protein、gene expression、differentiation、EMBRYONIC STEM、chromatin	The role of histone methyltransferase EZH2 in stem cell fate
161	acute myeloid、myeloid leukemia、stem cell、myelodysplastic syndrome、cell transplantation、patient、AML、hematopoietic stem、leukemia AML、treatment	Clinical application of hematopoietic stem cells
221	stem cell、retina、transplantation、cell、differentiation、retinal pigment、photoreceptor、pluripotent stem、MACULAR DEGENERATION、progenitor cell	Application of stem cells in retinal degenerative diseases

4.3　新兴研究主题拐点特征分析

利用一元三次函数，通过三次导函数的值进行判断，若 f(x)的三次导数不为 0，则该点为拐点。考虑到还在初期发展阶段的主题，拐点也可能在研究时段区间之外，本研究未给予拐点值阈值限制，比如限制在数据分析范围内。若该聚类中的样本个数少于 4 时，则无法进行函数拟合，拐点无法根据数据的模型拟合计算得出。获取拐点数据后，结果是否都合理需进一步分析。

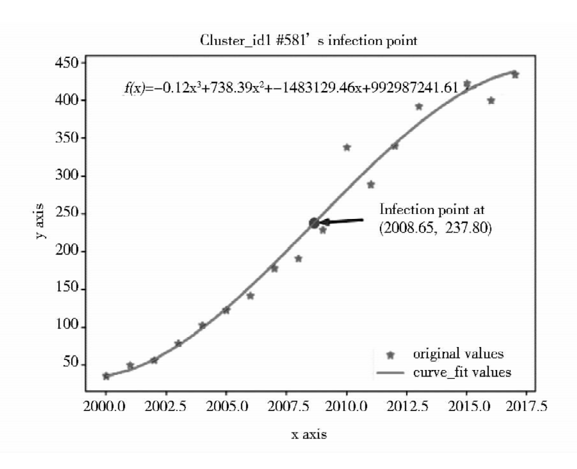
4.3.1　新兴研究主题的拐点类型划分

（1）拐点出现在分析时间区间内。19 个新兴研究主题的拐点出现在分析时间区间内，分别为#2、#60、#107、#142、#254、#353、#461、#581、#648、#710、#814、#867、#921、#965、#1046、#1095、#1142、#1199、#1460，这些主题在分析时间段内形成了最早有影响力的时间点。以#581 和#107 为例绘制带有拐点的演化曲线图，箭头指向的圆点为拐点时间（见图 5）。主题的拐点出现在时间段内的时间越晚越可能是新兴研究主题。#581在 2009 年左右出现拐点时间，而#107 在 2014 年左右出现拐点时间。

（2）拐点出现在分析时间区间之前。有两个新兴研究主题的拐点出现在时间区间之前，分别为#161、#221（见图 6）。结合领域专家分析以及对拟合曲线的分析，发现对该类型预测失误源自选取三次函数为模型的假设，事实上，这两个主题尚处于发展的早期阶段，未到达其拐点时间。

（3）拐点出现在分析时间区间之后。有两个新兴研究主题的拐点出现在时间区间之后，分别为#727 和#469（见图 7），可以预见到拐点会出现在未来一段时间。但发现对主题#469 未来拐点时间的预测可能存在问题，结合领域专家分析和数据特征，主题#469 确实处于快速发展阶段，但拐点的发文量可能要超过预测结果。因此，拐点出现在分析时间段之后的预测，也可能会出现失误，但拐点的预测结果依旧可以反映该主题当前还处于拐点出现前的发展阶段。





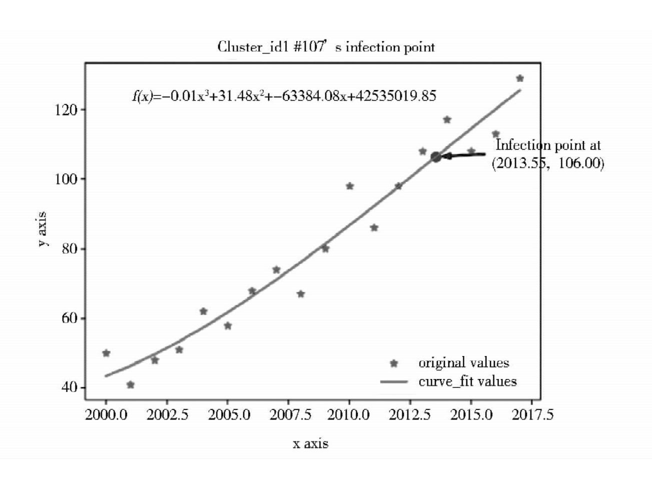
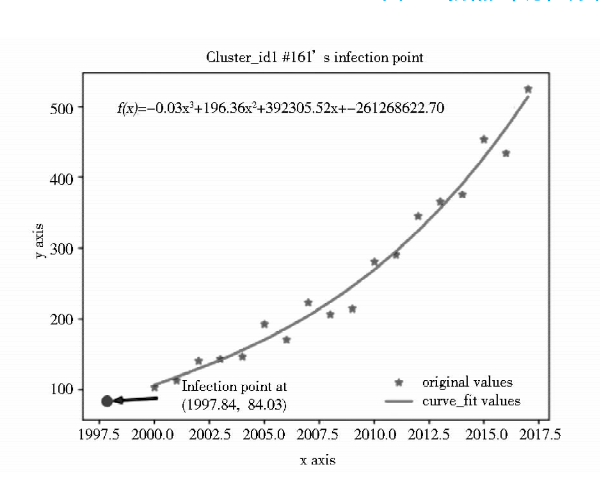


图 5　拐点出现在分析时间区间内



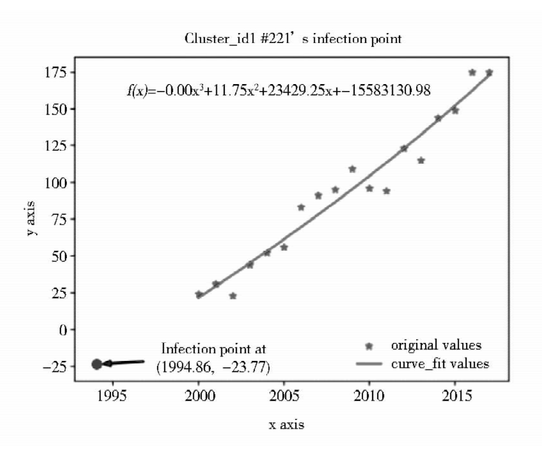
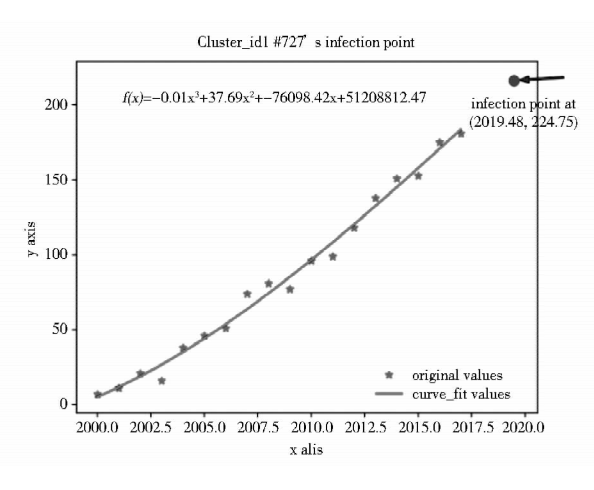


图 6　拐点出现在分析时间区间之前



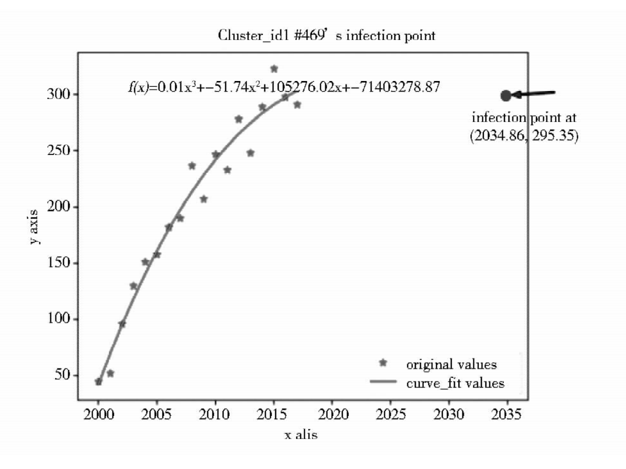
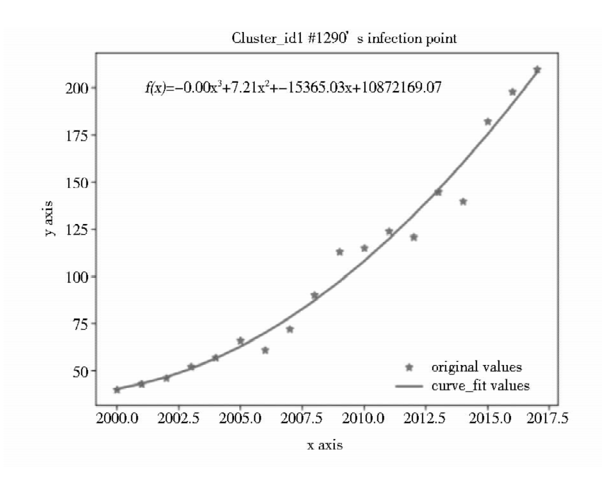


图 7　拐点出现在分析时间区间之后

（4）拐点未在分析时间区间内出现。有 3 个主题未能利用本文方法得到拐点时间，分别为#1290、#2276 和#261，其中#1290、#2276 的演化曲线见图 8。从图 8 中可以看出拐点未出现的主要原因是这些主题还处于新兴研究主题的早期发展阶段，尚不能做出拐点值的预测。

4.3.2　新兴研究主题的拐点时间与平均时间对比

26 个新兴研究主题的拐点与平均时间对比见表 3。表 3 中第 5 列"时间差_INF-AVG"为"平均时间_AVG"与"拐点时间_INF"的差值。"时间差_INF-AVG"的值为正，则"拐点时间_INF"早于"平均时间_AVG"；值为零，则"拐点时间_INF"等于"平均时间_



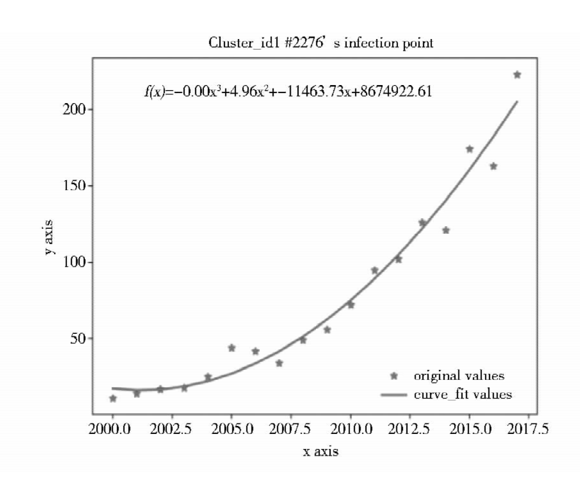


图 8　拐点未在分析时间区间出现

AVG”;值为负,则“拐点时间_INF”晚于“平均时间_AVG”。基于拐点值是对主题最早具备较大影响力的认识,那么对于新兴研究主题,“时间差_INF-AVG”为正值时,表明在统计时间段内,利用拐点时间可以比平均值更早识别该主题的影响力;“时间差_INF-AVG”为负值时表明该主题尚未到达具备一定影响力的阶段。

从表 3“时间差_INF-AVG”一列可以看出,除去没有拐点时间的主题#261、#1290 和#2276,剩余的 23 个主题中,有 6 个主题的“时间差_INF-AVG”为负值,3 个主题的“时间差_INF-AVG”为零,而有 17 个主题的“时间差_INF-AVG”为正值。由此证明在干细胞研究领域,拐点时间可以比平均值时间更早识别有影响力的主题。

新兴研究主题差值大于 5 年的有两个,分别为主题#221 和#161,拐点都在 1998 年之前,经核实属于数据拟合错误,这两个主题实际上还处于主题发展的早期阶段。

4.3.3　非新兴研究主题的拐点类型划分

与新兴研究主题一致,也将 28 个非新兴研究主题的拐点类型分为 4 类。在本实证分析中,对于非新兴研究主题,并没有拐点出现在分析时间区间之后的情况,这与非新兴研究主题自身特征相关,因为非新兴研究主题本身在筛选过程中就可能是近期增长缓慢的主题。

(1)拐点出现在分析时间区间内。22 个非新兴研究主题的拐点出现在分析时间段的区间内。以#899 和#23 为例(见图 9),这些主题在分析时间段内形成了最早有影响力的节点。拐点出现在时间段内的时间越晚,表明主题越是新兴研究主题。#899 在 2002 年左右出现拐点时间,而#23 在 2014 年左右才出现拐点时间。

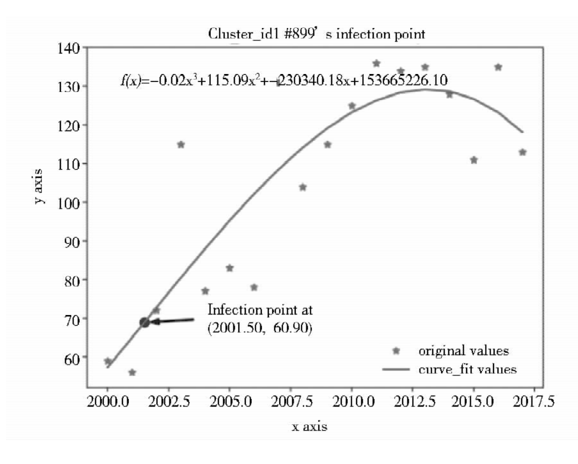
(2)拐点出现在时间区间之前。鉴于干细胞领域

表 3　新兴研究主题的拐点时间与平均时间对比

(单位/年)

主题序号	首次出现时间	拐点时间_INF	平均时间_AVG	时间差_INF-AVG
221	2000	1994	2009	15
161	2000	1998	2010	12
2	2001	2010	2014	4
254	2000	2009	2012	3
710	2000	2009	2012	3
1046	2000	2010	2012	2
1095	2000	2008	2010	2
142	2000	2010	2011	1
461	2000	2009	2010	1
581	2000	2009	2010	1
648	2000	2009	2010	1
814	2000	2009	2010	1
867	2000	2008	2009	1
1142	2000	2009	2010	1
921	2000	2010	2010	0
965	2000	2009	2009	0
1199	2000	2009	2009	0
60	2000	2011	2010	-1
353	2000	2012	2011	-1
1460	2000	2011	2010	-1
107	2000	2014	2008	-6
727	2000	2019	2010	-9
469	2000	2035	2009	-26
261	2000	-	2010	-
1290	2000	-	2010	-
2276	2000	-	2010	-

的研究始于 20 世纪 90 年代,大部分研究主题还处于新兴领域,早期信号不足,拐点可能在数据之外是合理的。本研究中,有 4 个非新兴研究主题的拐点出现在分析时间段之前,分别是#203、#285、#1623 和#1638。



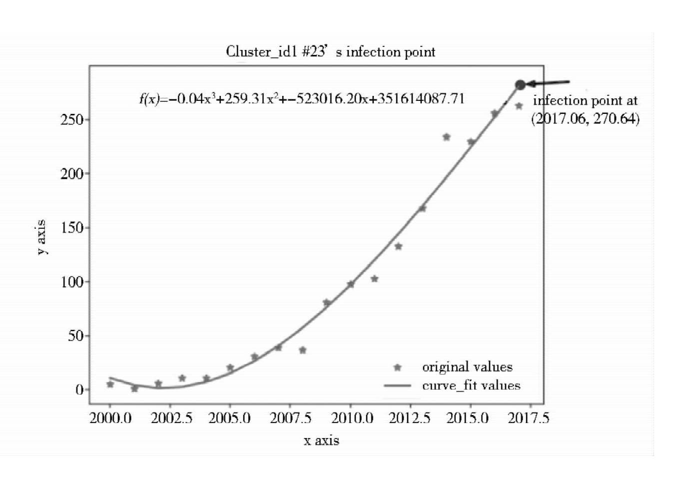
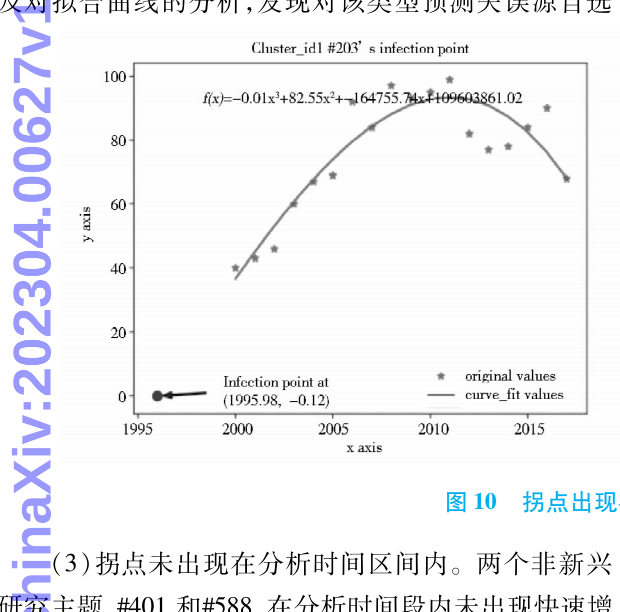


图 9 拐点出现在分析时间区间内

以#203 和#1638 为例(见图 10),结合领域专家分析以及对拟合曲线的分析,发现对该类型预测失误源自选取三次函数为模型的假设,事实上,这两个主题还处于自身发展的早期阶段,尚未到达拐点时间。



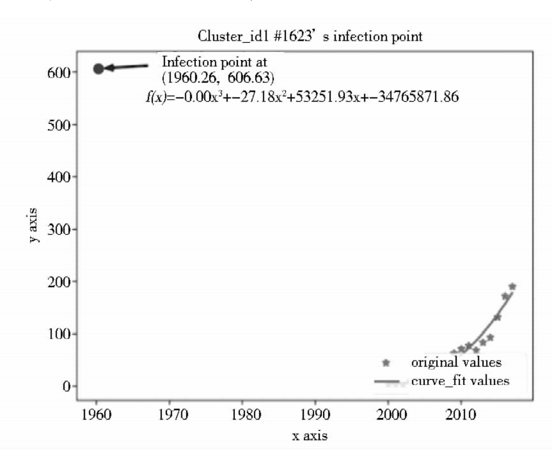
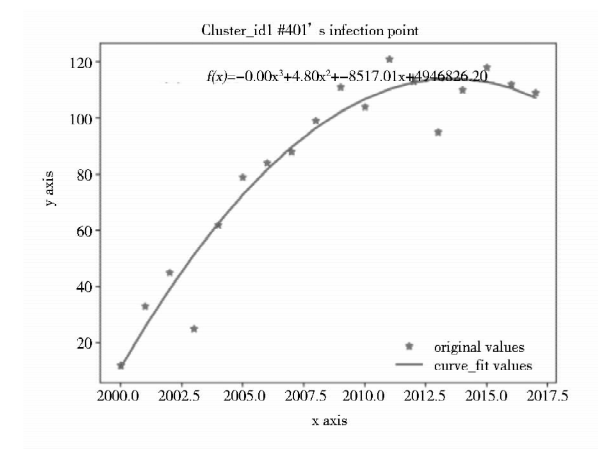


图 10 拐点出现在分析时间区间之前

(3)拐点未出现在分析时间区间内。两个非新兴研究主题,#401 和#588,在分析时间段内未出现快速增长趋势,并在近期有增长变缓的趋势,如图 11 所示:



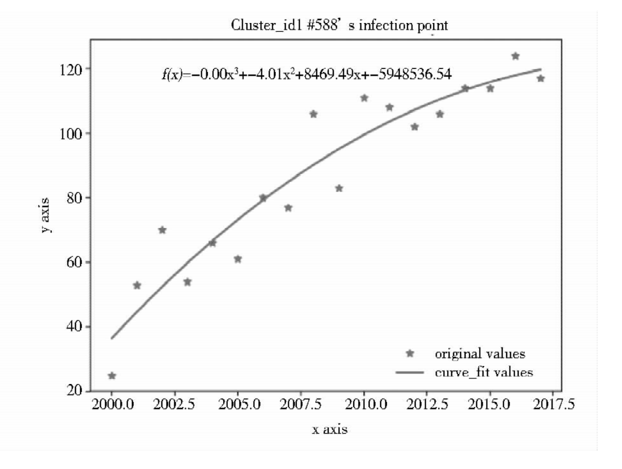


图 11 拐点未出现在分析时间区间

4.3.4 非新兴研究主题的拐点与平均时间对比

从表 4 中可以看出,第 5 列为 28 个非新兴研究主题"平均时间_AVG"与"拐点时间_INF"的差值"时间差_INF-AVG"。依据拐点值是对主题最早具备较大影

响力的认识，对于非新兴研究主题，“时间差_INF-AVG”的值为正值表明利用拐点时间可以比平均值更早识别该主题的影响力；“时间差_INF-AVG”的值为负值表明在统计时间段内，该主题尚未到达该主题发展的具备一定影响力的阶段。

从“时间差_INF-AVG”可以看出，除去没有拐点时间的主题#40 和#588，剩余 26 个主题中，有 6 个主题的“时间差_INF-AVG”为负值，3 个主题的“时间差_INF-AVG”为零，而有 17 个主题的“时间差_INF-AVG”为正。

相较于新兴研究主题，非新兴研究主题的正向差值更大，差值不小于 5 年的有 12 个，拐点时间均早于 2007 年，表明这些主题已经过了有影响力的早期阶段，已不属于新兴研究主题。

主题#1306、#23、#271 和#1265 的“时间差_INF-AVG”的值为负值，表明在统计时间段内，该主题尚未到达具备一定影响力的阶段，也尚未成为新兴研究主题。

因此，对于干细胞新兴研究主题，利用拐点时间可以比平均值时间提前识别有影响力的主题，且拐点对于新兴研究主题分析更为有效。

表 4　非新兴研究主题拐点时间、平均时间与首次出现时间对比　（单位/年）

主题序号	首次出现时间	拐点时间_INF	平均时间_AVG	时间差_INF-AVG
1623	2000	1960	2011	51
1638	2000	1995	2010	15
203	2000	1996	2009	13
285	2000	1998	2010	12
916	2000	2005	2014	9
1297	2000	2004	2013	9
899	2000	2002	2010	8
1267	2000	2002	2010	8
594	2000	2004	2011	7
168	2000	2005	2011	6
446	2000	2007	2012	5
850	2000	2005	2010	5
467	2000	2005	2009	4
215	2000	2008	2010	2
875	2000	2006	2008	2
2508	2000	2008	2010	2
810	2000	2008	2009	1
217	2000	2009	2009	0
659	2000	2011	2011	0
686	2000	2010	2010	0
451	2002	2013	2012	－1
2275	2000	2012	2008	－4
1306	2000	2016	2011	－5
23	2000	2017	2010	－7
271	2000	2016	2009	－7
1265	2000	2016	2009	－7
401	2000	–	2010	–
588	2000	–	2009	–

4.3.5　新兴研究主题与非新兴研究主题的拐点效用对比

通过对比发现：拐点对于新兴研究主题分析更为有效，存在数据拟合错误的情况更少。这主要是新兴研究主题呈现持续增长趋势，三次函数的拟合效果也更理想。对于非新兴研究主题而言，其主题发展趋势的模式除去在一段时间内呈现平稳增长或快速增长以外，可能存在更多发展模式，如趋于平稳或下降。因此，对于非新兴研究主题，其拐点时间分布更不确定，预测函数出错的几率也更高。总体来说，拐点对于新兴研究主题的识别效果更加稳定。

4.4　三种关键时间点的对比与应用策略

首次出现时间、平均时间、拐点时间分布优缺点见表 5。尽管拐点时间有其自身的重要性，但并非仅使用拐点值就可以完全实现主题的预测分析。研究主题会出现在路线图的不同时间段，而不同时间点的意义区别显著，创新路径中主题分布时间的确定需要综合 3 种时间类型。首次出现是该主题最早出现的信号，适合使用定量分析手段，但由于数据量稀少，难以得到关注；对于一个主题如果达到拐点则需要警戒，考虑是否需要及早布局；平均值则可以表征该主题是否已经成为一个研究热点。

表 5　三种时间分布优缺点

特征时间	优点	缺点
首次出现时间	首次出现时间必然会存在且容易获取，是主题出现的最早信号	影响力不足，难以引起注意
平均时间	平均值时间一定会存在且容易获取，可从数量特征上表现主题的影响力	由于主题论文发表的时间一般遵照指数增长模式，因此平均值一般往往容易滞后，难以评估这种影响力
拐点时间	拐点可以更好地表征主题发展的影响力变化。如果拐点时间早于平均时间，利用拐点更早预测主题的影响力；如果拐点晚于平均时间，说明该主题依旧处于早期发展阶段	可能不存在——可能来自于数据本身，也可能来自数据的拟合失误。不同拟合函数，也存在预测失误的情况

5　结语

本文主要侧重分析不同时间特征的主题发展差异，尤其是对早期识别新兴研究主题具有重要作用的生长曲线的拐点时间。笔者对比分析了首次出现时间、平均时间和拐点时间的差异，探讨新兴研究主题最早具有影响力的时间点。同时，该方法不仅适用于发文数量，也适用于其他反映主题生长曲线拐点分析，如引文曲线的拐点分析等。笔者将干细胞研究主题作为

实证研究对象，通过曲线拟合得到所识别的新兴研究主题的创新发展演化路径，并计算出研究发展路径中的拐点。研究成果有助于改进主题演化路径的准确性，提高主题演化路径在新兴研究主题的前瞻预判与布局的准确性。

本研究主要结论如下：首先，笔者依据拐点时间出现的区间范围将拐点类型分为4类：拐点出现在分析区间内、拐点出现在分析区间之前、拐点出现在分析区间之后、拐点未出现在分析区间内。通过对新兴研究主题与非新兴研究主题的拐点特征分析，大部分拐点出现在分析时间段的区间内，说明拐点对于新兴研究主题的预测具有较好的效用性。其次，对比分析新兴研究主题与非新兴研究主题的拐点时间与平均时间，发现利用拐点时间可以比平均值时间提前识别有影响力的主题。再次，对新兴研究主题与非新兴研究主题的拐点效用进行对比，发现拐点对于新兴研究主题分析更为稳定有效，存在拟合错误的情况更少；而对于非新兴研究主题而言，其拐点时间分布更为不确定，预测函数出错的几率也更高。最后，首次出现时间、平均时间和拐点时间在主题发展路径中意义区别显著，创新路径中主题分布时间需要综合应用3种时间类型。

本研究也存在几个缺陷：首先，笔者使用一元三次函数的假设前提无法提供严谨的论证，分析侧重从科学计量学中寻找依据，从科学计量的实践中分析意义与应用价值。本文尚未考虑其他数据分布类型，且并非所有曲线都遵照幂函数的方式，这也是导致拐点识别失误的主要原因。其次，三次函数无法有效识别多个拐点的情形，逻辑生长曲线也不适合所有主题的数据拟合，导致计量结果存在一定程度的失真。出现这种情况的原因在于每个主题都有自身发展特征，未来将尝试使用多个指标并结合具体领域的数据特征展开分析。再次，本文的分析数据源主要是科学论文，论文主要体现领域的基础研究成果，而专利可以体现一个领域的应用发展趋势。当前科技领域的发展往往是基础研究与应用研究协同并进，因此单一文献类型不能全面衡量领域的发展趋势。未来将尝试结合专利数据、产业经济数据等多源数据类型，形成对领域创新发展更全面的曲线模拟，拐点识别更准确。此外，未来将尝试对数据进行散点拟合的智能化，通过自适应过程识别每个主题分布曲线的最佳拟合函数。同时，将收集具体领域的全过程数据，尤其是主题的产生萌芽期，以期通过这些措施更准确地曲线拟合得到更准确的拐点时间。

参考文献：

[1] European Research Council. Identification of "Frontier Research" and "Emerging Research Areas" in research proposals: a bibliometric approach[EB/OL]. [2020-02-10]. https://erc.europa.eu/event/identification-frontier-research-and-emerging-research-areas-research-proposals-bibliometric.
[2] REISS T, VIGNOLA-GAGNÉ E, KUKK P, et al. ERACEP-emerging research areas and their coverage by ERC-supported projects [R]. Belgium: Technical Report European Research Council, 2013.
[3] Foresight and Understanding from Scientific Exposition (FUSE) [EB/OL]. [2020-02-10]. https://www.iarpa.gov/index.php/research-programs/fuse.
[4] WEI L, XU H Y, HU Z Y, et al. Multiple-pattern analysis and prediction of topic evolution path based on topic correlation: a case study of information science research[J]. Library and information service, 2016, 60(3):71-81.
[5] XU, H Y, YUE Z H, LUO R, et al. A study on the multidimensional scientometric indicators to detect the emerging topics[C] // Proceedings of the 17th international conference on scientometrics and informetrics (ISSI 2019). Italy:Sapienza University of Rome, 2019:2518-2519.
[6] YOUNG P. Technological growth curves:a competition of forecasting models[J]. Technological forecasting & social change, 1993, 44(4):375-389.
[7] KUCHARAVY D, GUIO R D. Problems of forecast[J]. ETRIA TRIZ future, 2005: 219-235.
[8] KUCHARAVY D, GUIO R D. Application of S-shaped curves [J]. Procedia engineering, 2011, 9: 559-572.
[9] KUCHARAVY D, GUIO R D. Logistic substitution model and technological forecasting[J]. Procedia engineering, 2011, 9:402-416.
[10] NGUIMKEU P. A simple selection test between the Gompertz and Logistic growth models[J]. Technological forecasting & social change, 2014, 88(Oct.):98-105.
[11] MEADE N, ISLAM T. Modelling and forecasting the diffusion of innovation-a 25-year review[J]. International journal of forecasting, 2006, 22(3):529-545.
[12] DAIM T U, RUEDA G, MARTIN H, et al. Forecasting emerging technologies: use of bibliometrics and patent analysis[J]. Technological forecasting and social change, 2006, 73(8):981-1012.
[13] KYEBAMBE M N, CHENG G, HUANG Y Q, et al. Forecasting emerging technologies: a supervised learning approach through patent analysis[J]. Technological forecasting and social change, 2017,125:236-244.
[14] 邵帅. 不确定性、经济预测与拐点[J]. 金融博览, 2019(4):35.
[15] 邱耕田. 发展的"拐点"：基于科学发展观视角的分析[J]. 学习与探索, 2009(3):26-31.

[16] 张学山. 经济学中的"拐点"与数学上的"拐点"[J]. 高等数学研究 2012,15(5):29－31.

[17] 解元元. 也谈拐点的定义[J]. 高等数学研究, 2003, 6(3):11－12.

[18] WASOW W. Linear turning point theory[M]. New York:Springer-Verlag, 1985.

[19] 郑步青, 邹红霞, 胡欣杰. 基于拐点的网络舆情预测研究[J]. 计算机科学, 2018, 45(S2):539－541,575.

[20] CAI F. Demographic transition, demographic dividend, and lewis turning point in China[J]. Economic research journal, 2010, 3(2):107－119.

[21] ROSTOW W, LEWIS W. The evolution of the international economic order[M]//International economics policies & their theoretical foundations. New York: Academic Press, 1982:15－37.

[22] WATANABE S. The lewisian turning point and international migration: the case of Japan[J]. Asian and Pacific migration journal: APMJ, 1994,3(1):119－147.

[23] EGLI H, STEGER T M. A dynamic model of the environmental kuznets curve: turning point and public policy[J]. Environmental and resource economics, 2007, 36(1):15－34.

[24] 华中炜, 郑伟. 库兹涅茨拐点决定中国未来经济走向[J]. 银行家, 2011(6):31－33.

[25] 魏君英, 吴亚平, 吴兆军. 基于库兹涅茨拐点的中国城乡居民收入差距与经济增长关系的实证研究[J]. 中国农业资源与区划, 2015, 36 (6):112－119.

[26] 朱小黄. 蒙格斯拐点及其学理[J]. 金融博览, 2019(2):37－39.

[27] 石陆魁, 秦志娇, 闫会强. 基于最小方差的股市拐点预测方法[J]. 计算机应用研究, 2017,34(11):3373－3378.

[28] CHRISTOU C, CUNADO J, GUPTA R, et al. Economic policy uncertainty and stock market returns in pacificrim countries: evidence based on a bayesian panel VAR model[J]. Journal of multinational financial management, 2017,40:92－102.

[29] PETROVSKA M, KRSTEVSKA A, NAUMOVSKI N. Forecasting macedonian business cycle turning points using qual Var model [J]. Nephron clinical practice, 2016, 5(3):61－78.

[30] SCHREIBER S, SOLDATENKOVA N. Anticipating business-cycle turning points in real time using density forecasts from a VAR[J]. Journal of macroeconomics, 2016, 47:166－187.

[31] ANDERSSON E, BOCK D, FRISEN M. Some statistical aspects of methods for detection of turning points in business cycles[J]. Journal of applied statistics, 2006, 33(3):257－278.

[32] ANDERSSON E, BOCK D, FRISÉN M. Detection of turning points in business cycles[J]. Journal of business cycle measurement and analysis, 2004(1):93－108.

[33] FUKUDA K. Forecasting growth cycle turning points using US and Japanese professional forecasters[J]. Empirical economics, 2009, 36(2):243－267.

[34] CANOVA F, CICCARELLI M. Forecasting and turning point predictions in a bayesian panel VAR model[J]. Journal of econometrics, 2004, 120(2): 327－359.

[35] BILLIO M, CASARIN R, RAVAZZOLO F, et al. Combination schemes for turning point predictions[J]. The quarterly review of economics and finance, 2012, 52(4):402－412.

[36] KOSKINEN L, OLLER L E. A classifying procedure for signalling turning points [J]. Journal of forecasting, 2004, 23 (3): 197－214.

[37] PELÀEZ R F. Ex ante forecasts of business-cycle turning points [J]. Empirical economics, 2007, 32:239－246.

[38] KULENDRAN N, WONG K K F. Determinants versus composite leading indicators in predicting turning points in growth cycle[J]. Journal of travel research, 2011, 50(4):417－430.

[39] JUDGE G. Turning-point errors and directional errors in forecasting: a note[J]. Journal of macroeconomics, 1985, 7(2):261－263.

[40] KULENDRAN N, WONG K K F. Predicting quarterly Hong Kong tourism demand growth rates, directional changes and turning points with composite leading indicators[J]. Tourism economics, 2009, 15(2):307－322.

[41] 邓华丽, 李修全. 基于混沌时间序列分析的股票价格拐点预测方法[J]. 统计与决策,2007(9):19－20.

[42] REISS J D. The analysis of chaotic time series[D]. Atlanta:Georgia Institute of Technology, 2001.

[43] LI X, DEN Z. Prediction of turning points for chaotic time series using ensemble ANN model[C] //2008 7th world congress on intelligent control and automation. Chongqing: IEEE, 2008:3459－3464.

[44] SCHARF J H. The position of the turning points in the growth curves during the intrauterine period in homo sapiens[J]. Acta anatomica, 1974, 88(4):520－540.

[45] 詹新宇. 广西产业结构服务化的拐点预测及对策探讨——基于增长曲线模型[J].广西社会科学,2011(4):24－28.

[46] HUBBERT M K. Energy resources: a report to the committee on natural resources of the national academy of Sciences——National research council[R]. Washington:National Academy of Sciences-National Research Council, 1962.

[47] 田红丽,李成群,闫会强.缠论和相似灰色模型的预测方法在股价拐点预测中的研究应用[J].计算机应用研究,2020,37(6):1666－1669,1678.

[48] JACKSON B P. A turning point for Europe's east[J]. Policy review, 2010(160):49－62.

[49] QU P P. Mobile communication service income prediction method based on grey buffer operator theory[J]. Grey systems: throry and application,2014,4(2):250－259.

[50] CLAVERIA O, MONTE E, TORRA S. Tourism demand forecasting with neural network models: different ways of treating information[J]. International journal of tourism research, 2015, 17(5):492－500.

[51] 胡悦. 金融市场中的神经网络拐点预测法[J]. 金融经济, 2017(18):81-82.
[52] 陈建华. 人工神经网络对企业发展拐点的预测研究[J]. 技术经济与管理研究, 2003(1):31-33.
[53] 马芳芳, 王京宝. 基于 RBF 神经网络的股票三类拐点的分析[J]. 科技信息(学术研究), 2008(6):211-214.
[54] GUO P C, WEI L, CHANG W X, et al. The contribution of logistics industry to economic growth based on logis model[C] // 2010 international conference of information science and management engineering. Shanxi: IEEE, 2010: 489-492.
[55] GOLDER P N, TELLIS G J. Will it ever fly? Modeling the takeoff of really new consumer durables[J]. Marketing science, 1997, 16(3): 256-270.
[56] ANDERSEN B. The hunt for S-shaped growth paths in technological innovation: a patent study[J]. Journal of volutionary economics, 1999,9(4): 487-526.
[57] XU H, WINNINK J, YUE Z, et al. Multidimensional scientometric indicators for the detection of emerging research topics [J]. Technological forecasting and social change, 2021, 163:120490.
[58] TRAAG V A, WALTMAN L, VANECK N J. From louvain to leiden: guaranteeing well-connected communities. [J]. Scientific reports, 2019, 9(1):1-12.
[59] 曹艺文, 许海云, 武华维, 等. 基于引文曲线拟合的新兴研究主题的突破性预测——以干细胞领域为例[J]. 图书情报工作, 2020, 64(5): 100-113.

作者贡献说明:

许海云:提出研究思路,设计研究框架,撰写论文;
张慧玲:文献调研与部分综述撰写;
武华维:数据处理与可视化,以及论文修改;
刘自强:部分数据处理。

Key Time-points of Emerging Research Topic on their Evolution Path

Xu Hainyun[1] Zhang Huiling[2] Wu Huawei[3] Liu Ziqiang[4]
[1] Business School, Shandong University of Technology, Zibo 255000
[2] Taiyuan Library, Taiyuan 030024
[3] Archives of Northwest Normal University, Lanzhou 730070
[4] School of Journalism and Communication, Nanjing Normal University, Nanjing 210023

Abstract: [**Purpose/significance**] To explore the different representations of the impact of different key time points on emerging research topics. [**Method/process**] Firstly, we summarized the application scenarios and acquisition methods of the current turning point time, and constructed the turning point identification method of emerging research topics on the innovation evolution path according to the growth mechanism and characteristics of network nodes in the knowledge diffusion. After that, the differences between "first appearance time" "average time" and "inflection point time" are compared and analysed, and explored the earliest point in time when emerging research topics have an impact. Finally, taking stem cell research topics as an empirical field, we analysed the different representational capabilities of different key time points on the influence of emerging research topics. [**Result/conclusion**] The turning point time can identify influential topics earlier than the "average time". "First appearance time" "average time" and "inflection point time" have significant differences in the topic evolution path. The determination of the distribution time of emerging research topics in the innovation path requires the synthesis of three different types of key time points.

Keywords: evolutionary path innovation path inflection point emerging research topic curve fitting